Canadian Journal of
Biomedical Research and Technology

Development of a Fast Screening Method for Osteoporosis Using Chest X-ray Images and Machine Learning

Open Access   **Research Article**

Ohta Y[1], Yamamoto K[2], Matsuzawa H[3], Kobayashi T[2], Ishida T[2*]

[1] MedCity21, Division of Premier Preventive Medicine, Osaka City University Hospital, Japan
[2] Department of Medical Physics and Engineering, Graduate School of Medicine, Osaka University, Japan
[3] Department of Radiology, Osaka University Hospital, Japan

*Address for Correspondence

Ishida T, Department of Medical Physics and Engineering, Graduate School of Medicine, Osaka University, Osaka University, Tel: +81668792577, Fax: +81668792577, Yamadaoka 1-7, Suita, Osaka 565-0871, Japan

Abstract

Objective: The international guidelines for the detection of osteoporosis and osteopenia recommend that patients undergo screening with the measurement of bone mineral density; however, the screening rate globally remains unsatisfactory. To resolve this concern, a technique for the identification of osteoporosis risk is required. Therefore, we aimed to introduce a fast-predictive method based on deep convolutional neural networks and detect objects indicating a high-risk of osteoporosis on chest x-ray images.

Materials and Methods: The proposed method was used to classify objects into high-risk and low-risk groups for osteoporosis using a deep convolutional neural network. We used chest radiographs and bone mineral densities measured by dual-energy x-ray absorptiometry of 320 female patients. Objects were annotated into the high-risk (n=160) or low-risk group (n=160) based on bone mineral densities. The dataset consisted of 24,073 regions of interest along the clavicles on chest radiographs. We trained the deep convolutional neural network with our training dataset. To prevent overfitting, 10-fold cross-validation was applied. This approach automatically generated the prediction for the classification.

Results: The classification accuracy, sensitivity, and specificity and area under the curve of the proposed method were 76.56%, 71.88%, 81.25%, and 0.7774, respectively. Generally, the misclassified results were caused by overlapping clavicles and ribs.

Conclusion: The deep convolutional neural network-based approach using chest radiographs was proven to be effective for detecting objects indicating a high osteoporosis risk. Our newly developed method may be helpful in improving screening rates.

Keywords: Bone mineral density; Deep convolutional neural network; Dual-energy x-ray absorptiometry; Osteopenia; Osteoporosis; Screening

Abbreviations: AUC= Area under the curve; BMD= Bone mineral density; CXI= Chest x-ray image; DCNN= Deep convolutional neural networks; DXA= Dual-energy x-ray absorptiometry; ROC= Receiver operating characteristic; ROI= Region-of-interest; YAM= Young adult mean

Introduction

Osteoporosis is characterized by an imperceptible and painless reduction in bone mineral density (BMD) and deterioration of the bone microarchitecture, which leads to an increased risk of patients becoming bedridden due to hip, wrist, and spine fractures [1]. Currently, more than 200 million people worldwide are estimated to be affected by osteoporosis, and the disease is considered a serious public health concern [2,3]. Statistically, it is well known that the individuals who have the highest risk of developing osteoporosis are elderly postmenopausal women; however, it is a preventable and treatable disease [3]. Thus, it is recommended worldwide that elderly women should pay attention to and increase their awareness of their own bone health for early detection of decreased BMD symptoms which include not only osteoporosis but also osteopenia. In general, dual-energy x-ray absorptiometry (DXA) is the most common method for accurately measuring BMD [4]. There are several advantages of measuring BMD using DXA; particularly, the data acquisition process is simple, fast, and painless. However, screening rates still remain unsatisfactory [5-7]. In clinical practice, Dell and Greene [8] reported that only 27% of eligible women aged 66-70 years old underwent DXA screening. To improve this situation, medical organizations need a simple method for identifying individuals with lower BMDs estimated BMD values on hip and chest x-ray images (CXIs) [9]. In their study, the cortical thicknesses of the femur shaft and clavicles on x-ray images were used for BMD estimations. Although this is a high-performance method for BMD estimations and osteoporosis predictions and CXIs are most commonly captured during health checkups, capturing hip radiographs is rare in daily life. To identify individuals with low BMDs in

order to recommend osteoporosis examinations, it is necessary to be able to use only easily accessible sources of information, such as CXIs. Furthermore, Kumar's method requires many manual measurements for extracting relevant features. In recent years, deep convolutional neural networks (DCNNs) have been widely used for computer-aided disease detection and classification in the field of medical image processing and have exhibited high performance [10,11]. DCNNs can automatically extract features by using convolution layers. Each layer has filters that are automatically generated on learning large image datasets with accurate labels, enabling the extraction of the features required for target identification. In this study, we aimed to introduce a method for detecting objects indicating lower BMDs on CXIs by using a DCNN.

Materials and Methods

Patients

We obtained CXIs and BMD values using DXA in 320 female patients (aged 27-82 years old) between 2014 and 2017. Patients with lumbar compression fractures and femoral neck fractures were excluded because of inaccurate BMD values. All the patients underwent chest radiography and DXA within 2 weeks. Standard digital chest posterior-to-anterior view radiographs were captured using three types of digital x-ray machines: Dr. Beneo (Fujifilm, Tokyo, Japan), Digital

Diagnost 2.1 (Philips Healthcare, Best, Netherland), and Definium 8000 (GE Healthcare, Chalfont St. Giles, UK). The following parameters were used: x-ray tube voltage, 120 kV; tube current exposure time product, 2-3 mAs; and focus-to-film distance, 180 cm. In this study, well-trained radiological technologists measured the BMDs using a DXA scanner, QDR-4500 (Hologic, Waltham, MA, USA), and the standard protocol. In each patient, BMDs were measured at L1-L4 (the first to the fourth lumbar vertebrae), L2-L4 (the second to the fourth lumbar vertebrae), and the femoral neck. Subsequently, the lowest value among these three measured values was recorded as the patient's BMD.

Pre-Processing

The resolution of the CXIs was 143 to 200 μm. First, we created and annotated the region-of-interest (ROI) images in order to train the DCNN. Following the Japan Osteoporosis Society Guideline, the objects that exhibited young adult mean (YAM) values of less than 80 % were regarded as necessary for detailed examinations [12]. YAM value is defined as the average BMD of healthy adults aged 20 to 44 years with their bone densities maintained at the maximum level, and the %YAM value is the ratio of measured BMD to YAM. We annotated each CXI into a high-risk (%YAM<80) group and low-risk (%YAM≥80) group. As shown in Figure 1,
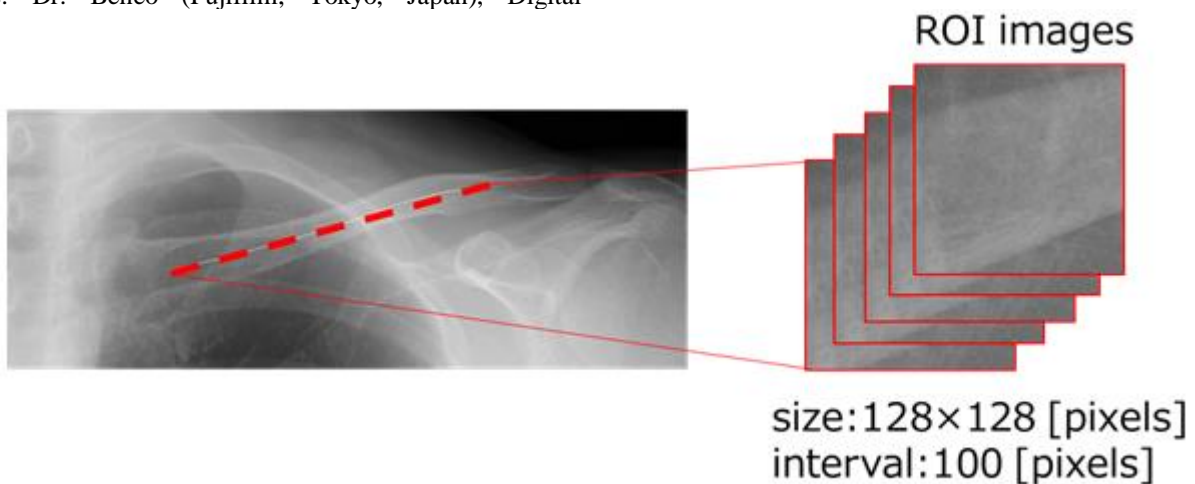
Figure 1: Overview on setting centerlines and extracting ROIs from CXIs. The centerlines were manually set. Subsequently, ROIs were obtained from every 100 pixels along the centerlines.

Abbreviations:

ROI= Region of interest
CXI= Chest x-ray image
The relevant ROIs (128×128 pixels) were acquired by setting ROIs within every 100 pixels along the centerlines on both the clavicles on the CXIs. By using this method, we facilitated the acquisition of a large number of ROIs for training the DCNN using a limited number of objects. The centerlines were manually set within the clear and visible range of the cortical bone on the clavicles. In total, 24,073 ROIs were obtained from

320 objects. To demonstrate the versatility and validity of our approach by using the cross-validation method, 320 objects in this study were randomly divided into 10 groups. MATLAB 2019b (MathWorks, Natick, MA, USA) and ImageJ 1.51p (National Institutes of Health, Bethesda, MD, USA) were used for image processing.

DCNN and Transfer Learning

To classify the ROIs extracted from the objects into the high-risk and low-risk groups based on the risk of osteoporosis, we conducted transfer learning which is one of the abilities of a

DCNN. In transfer learning, the pre-trained weights necessary for calculations in a DCNN are applied as the starting point for learning a new task. Compared with learning from the initial state, the advantage of transfer learning lies in the fact that it is possible for a DCNN to learn image features within a small image in a short duration of time [13]. Moreover, the effectiveness of the process has been reported by previous studies [14,15]. We applied ResNet50 architectures that were pre-trained on approximately 1.2 million training images and 50,000 validation images (1,000 object categories) from the ImageNet Large-Scale Visual Recognition Challenge [16,17]. To identify targets, DCNNs contain some convolution layers for the extraction of features from targets on images. In the transfer learning process conducted in this study, only the final layers were replaced in order for the target tasks to be adapted. The experiments were conducted in the MATLAB 2019b environment on Intel® CoreTM i7-9700K CPU (Intel Santa Clara, CA, USA) and GeForce GTX 1060 6GB (NVIDIA Corporation, Santa Clara, CA, USA). The DCNN was trained in 30 epochs. We used a stochastic gradient descent with a momentum of 0.9. The learning rate was fixed at 0.0001. In this study, every ROI obtained from a single case was classified into the high-risk or low-risk group. We adapted the class that had a larger number of classified ROIs in order to determine the final class of each object. The determined classes were used to calculate the accuracy, sensitivity, and specificity of the classification.

Analysis

Statistical analyses were performed using MATLAB 2019b. The mean of each index was calculated, and the confidence intervals for the means were determined using a normal distribution. To evaluate the performance of the trained DCNN, the classification accuracy, sensitivity, and specificity were calculated. Additionally, receiver operating characteristic (ROC) curves were generated, and the area under the curve (AUC) was determined. In the ROC analysis, the closer the AUC value was to 1, the better the screening performance results were in identifying objects indicating the risk of osteoporosis. AUC values between 0.5 and 0.7 indicate a poor performance, those between 0.7 and 0.9 indicate a moderate performance, and those between 0.9 and 1 indicate that the tool has high predictability [18]. In addition, the average menopausal age in Japan is approximately 50 years [19], and based on this age, the objects derived from the patients were divided into two groups: a group under 49 years old and a group over 50 years old. We evaluated the performance of the proposed method in each group.

Consistency of the Experiment

We investigated the consistency of the classification accuracy obtained through the trained DCNN under various imaging conditions. On assuming the possible variety of available imaging devices and image settings, we created each test image dataset by including noise, blur, contrast change, and compression, respectively. The additional noise parameter was Gaussian noise at a standard deviation of 2.0. The images were blurred using a filter that replaced each pixel with the average of its 3×3 pixel neighborhood. In order to change the contrast, the window width was set to 1.5 times the width of the original images. For image compression, the matrix size of each image was set to 64×64 pixels. Furthermore, the process of obtaining ROIs from the clavicles on CXIs was manually practiced. To investigate the effect of this process on classification accuracy, an observer group that consisted of a doctor and five radiological technologists carried out the process and created each test image dataset. Each test image dataset was classified by using the trained DCNN that exhibited the highest classification accuracy in 10-fold cross-validation.

Results

	All Objects (n=320)			High-risk (n=160)			Low-risk (n=160)		
	Mean	SD	95%CI	Mean	SD	95%CI	Mean	SD	95%CI
Age (years)	60.27	12.63	58.88-61.65	64.23	11.43	62.46-66.00	56.3	12.53	54.36-58.24
Weight (kg)	54.63	12.11	53.30-55.95	49.75	8.62	48.4-51.09	59.5	13.09	57.47-61.53
Height (cm)	156.1	6.06	155.4-156.7	154.9	5.76	153.9-155.8	157.2	6.13	156.3-158.2
BMI	22.44	4.79	21.91-22.96	20.76	3.56	20.20-21.31	24.12	5.25	23.30-24.93
BMD from L1 to L4	0.87	0.17	0.85 to 0.88	0.76	0.13	0.73 to 0.78	0.98	0.12	0.96-0.99
BMD from L2 to L4	0.89	0.18	0.87 to 0.91	0.77	0.14	0.75 to 0.79	1.01	0.12	0.99-1.03
BMD from femoral neck	0.64	0.13	0.62 to 0.65	0.54	0.08	0.53 to 0.56	24.12	5.25	0.72-0.75

Table 1: Characteristics of the study objects

Canadian Journal of
Biomedical Research and Technology

Abbreviations: BMI=Body mass index
BMD=bone mineral density
L1 to L4=The first to the fourth lumbar vertebrae
L2 to L4=The second to the fourth lumbar vertebrae
CI=Confidence interval
SD=Standard deviation.

(Table 1) presents the average ± standard deviation of age, weight, height, body mass index, and each BMD value of the included cases.

Group	Number of objects			Acc.	Sen.	Spe.	AUC
	All	%YAM<80	%YAM≥80				
All objects	320	160	160	76.56%	71.88%	81.25%	0.7774
Under 49 years old	64	17	47	81.25%	58.82%	89.36%	0.7144
Over 50 years old	256	143	113	75.39%	73.43%	77.88%	0.772

Table 2: Performances of proposed method for identifying

objects which show lower BMDs.

Abbreviations: %YAM=percent young adult mean; Acc.=Accuracy; Sen.=Sensitivity; Spe.=Specificity; AUC=Area under the curve

(Table 2) presents the classification results obtained using the DCNN. Using the proposed method, the cases were classified into the high-risk group (%YAM<80) and low-risk group (%YAM≥80) in order to identify the risk of osteoporosis. The classification accuracy, sensitivity, and specificity for all the objects were 76.56%, 71.88%, and 81.25%, respectively. All the networks' training accuracies converged within 30 epochs. In the analysis based on age, the classification accuracy, sensitivity, and specificity in the group of subjects under 49 years old (n=64) were 81.25%, 58.82%, and 89.36%, respectively, and those in the group of subjects over 50 years old (n=256) were 75.39%, 73.43%, and 77.88%, respectively.
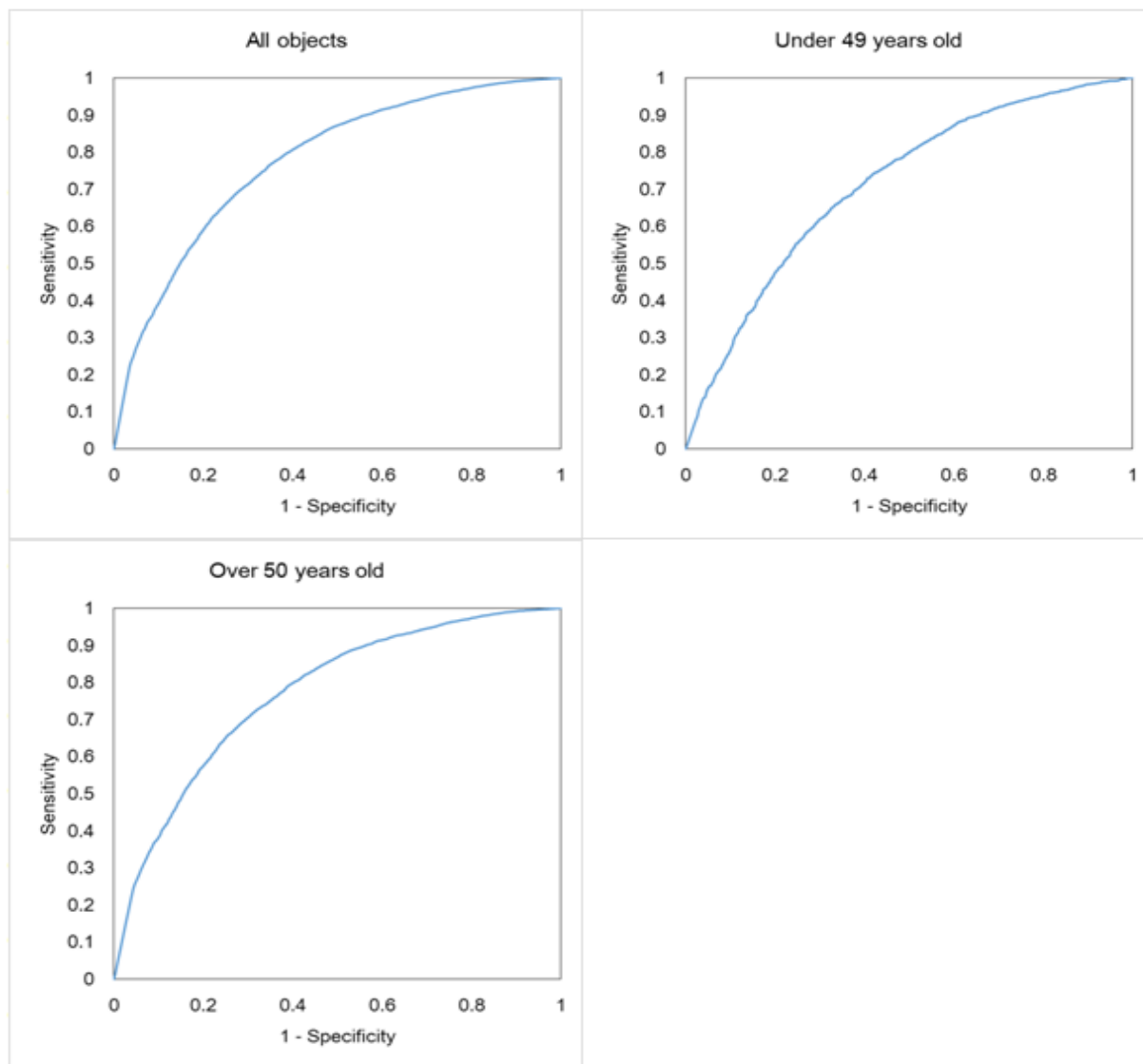
Figure 2: ROC curves of the proposed method for identifying objects that exhibit lower BMDs.

Abbreviations: ROC= Receiver Operating Characteristics; BMD= Bone Mineral Density

(Figure 2) presents the ROCs for each test dataset. The AUC values of the trained DCNN in the three different datasets comprising of all the objects, objects from the group under 49 years old, and objects from the group over 50 years old were 0.7774, 0.7144, and 0.7720, respectively. To evaluate the influence of the manual ROI acquisition process on classification accuracy, we used the training and test datasets, which exhibited the highest classification accuracy in 10-fold cross validations.

	Author	Observer 1	Observer 2	Observer 3	Observer 4	Observer 5	Observer 6
Specialty	-	MD	RT	RT	RT	RT	RT
Experience (years)	-	more than ten	more than ten	more than ten	three to five	less than one	less than one
Number of ROIs	1913	2158	1565	2633	1517	1402	1139
Overall Accuracy (%)	96.88	96.88	100	93.75	96.88	81.25	96.88

Table 3: Observers'specialty, experience, number of their datasets and result.

Abbreviations: ROI= Region of Interest; RT= Radiological Technologist; MD-Medical Doctor

(Table 3) shows the classification accuracies and number of ROIs obtained from each test dataset by the observers. We acquired 1,913 ROI images from the test dataset. The observer group consisted of six individuals: a doctor and two radiological technologists with more than 10 years of experience, a radiological technologist with 3 to 5 years of experience, and two radiological technologists with less than 1 year of experience. The average classification accuracy in each test dataset generated by the observers was 94.27±6.68%, and the average number of acquired ROI images was 1,735±552.8. The classification accuracies in each test dataset were high regardless of the observers' experience and number of ROIs on the images. The classification accuracies in each test dataset with added noise, blur, contrast change, and compression were 62.5%, 65.6%, 93.8%, and 62.5%, respectively. The classification results in the dataset in which the contrast change was added were similar to the results of the original dataset. However, the classification accuracies in the datasets with the other added processing elements were low.

Discussion

DCNNs are frequently used to detect features and perform classifications on images *[14,15]*. In transfer learning, the pre-trained weights that are needed by DCNNs for feature extraction are applied as a starting point for the network to learn a new task. Compared with learning from the primary state, the advantage of transfer learning is that it minimizes the load and working time of the analytical device *[13]*. Generally, because of the increasing calculation costs, the deeper the layers of networks are, the more difficult their training is. The Resnet architecture used in this study has a residual learning framework and makes deeper network training easier than with other previously used networks. Because of this framework, Resnet can produce high accuracy *[16]*. The AUC of the proposed method was 0.7774 and indicated a moderate screening performance, which helped us to be able to state that only CXIs were required for recognizing lower BMDs.

Therefore, the use of this method could be proposed as an initial and easy screening technique. In the age-based evaluation, the sensitivity in the group under 49 years old was low. It is possible that the DCNN was not fully able to learn the features within objects from subjects under 49 years old since the used dataset mostly consisted of objects from subjects over 50 years old. To improve the performance, we need to train the DCNN with various datasets. To visualize the region on an image that the DCNN used to make a decision, we created heat maps by using class activation mapping (CAM) *[20]*.
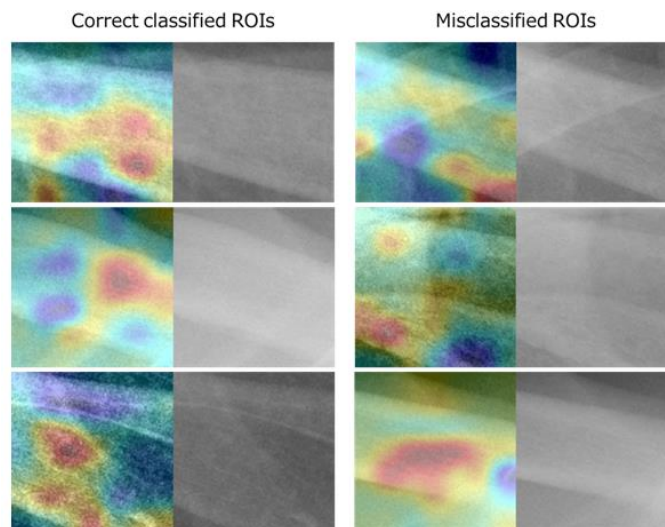
Figure 3: Examples of heat maps depicting correctly classified and misclassified ROIs: the first column indicates the corresponding heat maps, and the second column indicates the ROIs; the redder the color of the area, the more strongly the classification by the DCNN was affected.

Abbreviations: ROI= Region of interest; DCNN= Deep Convolutional Neural Network

(Figure 3) shows an example of correctly classified ROIs and misclassified ROIs based on our practice in order to visualize

the classification process of the trained DCNN. The redder the color of the area, the more strongly the classification of the DCNN was affected. In the correctly classified ROIs, the red areas were mainly distributed within the texture inside the clavicle and cortical bone. On the other hand, in the misclassified ROIs, the red areas were distributed in the areas that exhibited high pixel values due to the overlap of ribs, muscles, and fat. Thus, the DCNN's use of structures other than the clavicles for classification may be the cause of misclassification. This possibility is one of the reasons for misclassification. In this study, an enormous number and various types of ROIs were obtained from each case. These obtained ROIs included ROIs that were not suitable for classification because they contained other structures, such as the ribs and scapulae, which overlapped the clavicles. The exclusion of ROIs that could lead to misclassifications is an issue for us to resolve in the future. To evaluate the effect of the manual ROI acquisition process on classification accuracies, we classified each test dataset consisting of ROIs manually acquired by the observers. The original classification accuracy was 96.88%. In contrast, the average classification accuracy in the test dataset acquired by the observers was 94.27±6.68%. Regardless of the observers' specialties and experience levels and the number of ROIs, the proposed method exhibited high classification accuracies. It is considered that the process involving ROIs being acquired manually could not significantly affect the classification accuracy. On analyzing the classifications in the test dataset with various added image processing conditions, we found that these conditions affect classification accuracies. Similar results have been reported in previous studies *[21]*. On the other hand, the proposed method was uniformly able to classify radiographs captured using three different types of x-ray equipment. Therefore, we expect the proposed method to be widely used for classifications on CXIs captured using standard protocols. It is well known that postmenopausal women should routinely be screened for osteoporosis using DXA. In particular, the screening interval should be reduced in the case of individuals with lower BMDs [22]. We frequently need to confirm whether an object's BMD is lower than the criterion without excessive burdens. Because our proposed method is only based on classifications on a CXI, it does not require any special operations or special terminal devices. The proposed method has great accessibility for recognizing lower BMDs, and this ability may contribute to improving osteoporosis screening rates globally.

Conclusion

We have developed an effective method for detecting objects with lower BMDs on CXIs using a DCNN-based approach. Since our proposed method involves simplified processing only, it contributes to improved osteoporosis screening rates. There are some limitations to this study. Firstly, the sample size was insufficient, especially, of younger patients. Therefore, the classification accuracies in the younger group were poor. To utilize the proposed method in screening people of all ages, we need to train DCNNs using various objects. Secondly, the screening performance of our proposed method demonstrated

moderate performance. It is necessary to further improve the performance before implementing it in practice. Future research in this field should further investigate the ROI acquisition and pre-processing methods.

Acknowledgment

Compliance with Ethical Standards

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethical Approval

All of the procedures performed in the study involving human participants were in accordance with the ethical standards of the 1964 Helsinki declaration and its later amendments or comparable ethical standards, and the Institutional Review Board approved the study.

Informed Consent

Formal consent is not required for this type of study, and this article does not contain patient data.

References

1. Hellekson KL (2002) NIH releases statement on osteoporosis prevention, diagnosis, and therapy. Am Fam Physician 66:161-162.
2. Cooper C, Campion G, Melton LJ 3rd (1992) Hip fractures in the elderly: A world-wide projection. Osteoporosis Int 2: 285-289.
3. Sözen T, Özışık L, Başaran NC (2017) An overview and management of osteoporosis. Eur J Rheumatol 4: 46-56.
4. Karunanithi R, Ganesan S, Panicker TMR, Korath MP, Jagadeesan K (2007) Assessment of bone mineral density by DXA and the trabecular microarchitecture of the calcaneum by texture analysis in pre-and postmenopausal women in the evaluation of osteoporosis. J Med Phys 32:161-168.
5. Mithal A, Bansal B, Kyer CS, Ebeling P (2014) The Asia-Pacific regional audit-epidemiology, costs, and burden of osteoporosis in India 2013: A report of international osteoporosis foundation. Indian J Endocrinol Metab 18: 449-454.
6. Kanis JA, Borgström F, Compston J, Dreinhöfer K, Nolte E, et al., (2013) SCOPE: A scorecard for osteoporosis in Europe. Arch Osteoporos 8: 144.
7. Zanchetta J, Campusano C, Muzzi B, Terront A, Cerdas S, et al., (2012) The Latin America regional audit: Epidemiology, costs&burden of osteoporosis in 2012. International Osteoporosis Foundation: 45-52.
8. Dell R, Greene D (2010) Is osteoporosis disease

management cost effective? Curr Osteoporos Rep 8: 49-55.

9. Kumar DA, Anburajan M (2014) The role of hip and chest radiographs in osteoporotic evaluation among south Indian women population: A comparative scenario with DXA. J Endocrinol Invest 37: 429-440.

10. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, et al., (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542: 115-118.

11. Ardila D, Kiraly AP, Bharadwaj S, Choi B, Reicher JR, et al., (2019) End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 25: 954-961.

12. Soen S, Fukunaga M, Sugimoto T, Sone T, Fujiwara S, et al., (2013) Diagnostic criteria for primary osteoporosis: Year 2012 revision. J Bone Miner Metab 31: 247-257.

13. Tajbakhsh N, Shin JY, Gurudu SR, Hurst T, Kendall CB, et al., (2016) Convolutional neural networks for medical image analysis: full training or fine tuning? IEEE Trans Med Imaging 35: 1299-1312.

14. Näppi JJ, Hironaka T, Regge D, Yoshida H (2016) Deep transfer learning of virtual endoluminal views for the detection of polyps in CT colonography. In Medical Imaging 2016: Computer-Aided Diagnosis. Int Soc Optics Photo 97852B.

15. da Nóbrega RVM, Peixoto SA, da Silva SPP, Filho PPR (2018) Lung nodule classification via deep transfer learning in CT lung images. Proceedings of IEEE 31st International Symposium on Computer-Based Medical Systems 2018; Karlstad, Sweden 244-249.

16. Kaiming H, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. Proceedings of IEEE conference on computer vision and pattern recognition 770-778.

17. (2016) ImageNet: Stanford Vision Lab, Stanford University, Princeton University.

18. Akobeng AK (2007) Understanding diagnostic tests 3: Receiver operating characteristic curves. Acta Paediatr. 96: 644-647

19. Tamada T, Iwasaki H (1995) Age at natural menopause in Japanese women. Acta Obstet Gynaecol Jpn 47: 947-952.

20. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A (2016) Learning deep features for discriminative localization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 2921-2929.

21. Dodge S, Lina K (2016) Understanding how image quality affects deep neural networks. Proceedings at the Eighth International Conference on Quality of Multimedia Experience (QoMEX): IEEE 1-6.

22. Gourlay ML, Fine JP, Preisser JS (2012) Bone-density testing interval and transition to osteoporosis in older women. N Engl J Med 366: 225-233.

Assets of Publishing with us

Global archiving of articles

Immediate, unrestricted

online access Rigorous Peer

Review Process Authors

Retain Copyrights

https://www.biomedress.com

Submission Link: https://biomedress.com/online-submission.php